# SSB ban - Census validation

*Sanjay Basu*

*5/30/2019*

## Compare NHANES to Census

for face validity

```
census_api_key("502c9ebce6cbb78e55b5799866a5771ad6411885", install=T, overwrite=T)
```

```
## Your original .Renviron will be backed up and stored in your R HOME directory if needed.

## Your API key has been stored in your .Renviron and can be accessed by Sys.getenv("CENSUS_API_KEY").
## To use now, restart R or run `readRenviron("~/.Renviron")`

## [1] "502c9ebce6cbb78e55b5799866a5771ad6411885"
```

```
readRenviron("~/.Renviron")
v17 <- load_variables(2017, "acs5", cache = TRUE)
```

## Extract relevant ACS variables

civilian employed population by age and sex

```
emp <- get_acs(geography = "us",
               variables = c(m_16_19 = "B23001_007",
                             m_20_21 = "B23001_014",
                             m_22_24 = "B23001_021",
                             m_25_29 = "B23001_028",
                             m_30_34 = "B23001_035",
                             m_35_44 = "B23001_042",
                             m_45_54 = "B23001_049",
                             m_55_59 = "B23001_056",
                             m_60_61 = "B23001_063",
                             m_62_64 = "B23001_070",
                             m_65_69 = "B23001_075",
                             m_70_74 = "B23001_080",
                             m_75_99 = "B23001_085",
                             f_16_19 = "B23001_093",
                             f_20_21 = "B23001_100",
                             f_22_24 = "B23001_107",
                             f_25_29 = "B23001_114",
                             f_30_34 = "B23001_121",
                             f_35_44 = "B23001_128",
                             f_45_54 = "B23001_135",
                             f_55_59 = "B23001_142",
                             f_60_61 = "B23001_149",
                             f_62_64 = "B23001_156",
```

```
                               f_65_69 = "B23001_161",
                               f_70_74 = "B23001_166",
                               f_75_99 = "B23001_171"))
```

## Getting data from the 2013-2017 5-year ACS

```
emp$prop = emp$estimate/sum(emp$estimate)
```

## Compare to NHANES

T-test for difference, with comp matrix as final results by age/sex category

```
impnhanes = read_csv("imp_nhanes.csv")
```

```
## Parsed with column specification:
## cols(
##   .default = col_double(),
##   female = col_logical(),
##   white = col_logical(),
##   black = col_logical(),
##   asian = col_logical(),
##   other = col_logical(),
##   hisp = col_logical()
## )
```

```
## See spec(...) for full column specifications.
```

```
tally = impnhanes %>%
  mutate(m_16_19 = (female==0)&(age>=16)&(age<=19),
         m_20_21 = (female==0)&(age>=20)&(age<=21),
         m_22_24 = (female==0)&(age>=22)&(age<=24),
         m_25_29 = (female==0)&(age>=25)&(age<=29),
         m_30_34 = (female==0)&(age>=30)&(age<=34),
         m_35_44 = (female==0)&(age>=35)&(age<=44),
         m_45_54 = (female==0)&(age>=45)&(age<=54),
         m_55_59 = (female==0)&(age>=55)&(age<=59),
         m_60_61 = (female==0)&(age>=60)&(age<=61),
         m_62_64 = (female==0)&(age>=62)&(age<=64),
         m_65_69 = (female==0)&(age>=65)&(age<=69),
         m_70_74 = (female==0)&(age>=70)&(age<=74),
         m_75_99 = (female==0)&(age>=75)&(age<=99),
         f_16_19 = (female==1)&(age>=16)&(age<=19),
         f_20_21 = (female==1)&(age>=20)&(age<=21),
         f_22_24 = (female==1)&(age>=22)&(age<=24),
         f_25_29 = (female==1)&(age>=25)&(age<=29),
         f_30_34 = (female==1)&(age>=30)&(age<=34),
         f_35_44 = (female==1)&(age>=35)&(age<=44),
         f_45_54 = (female==1)&(age>=45)&(age<=54),
         f_55_59 = (female==1)&(age>=55)&(age<=59),
         f_60_61 = (female==1)&(age>=60)&(age<=61),
```

```r
        f_62_64 = (female==1)&(age>=62)&(age<=64),
        f_65_69 = (female==1)&(age>=65)&(age<=69),
        f_70_74 = (female==1)&(age>=70)&(age<=74),
        f_75_99 = (female==1)&(age>=75)&(age<=99)) %>%
  select(matches("^m_|f_"))


cols <- sapply(tally, is.logical)
tally[,cols] <- lapply(tally[,cols], as.numeric)
nhanes = colSums(tally)/dim(tally)[1]
census = (cbind(emp$variable, emp$prop))
comp = (cbind(census, nhanes))
print(comp)
```

```
##                                                    nhanes
## m_16_19 "m_16_19" "0.0160563174437255"  "0.031922525107604"
## m_20_21 "m_20_21" "0.017847549154738"   "0.0170970827355332"
## m_22_24 "m_22_24" "0.0309822235734176"  "0.0341941654710665"
## m_25_29 "m_25_29" "0.058857112521175"   "0.0575083692013391"
## m_30_34 "m_30_34" "0.0591508126887689"  "0.0633668101386896"
## m_35_44 "m_35_44" "0.112358186049704"   "0.109158297465328"
## m_45_54 "m_45_54" "0.114359897015365"   "0.104854136776662"
## m_55_59 "m_55_59" "0.0508419153585613"  "0.0449545671927307"
## m_60_61 "m_60_61" "0.0170298022568717"  "0.0196078431372549"
## m_62_64 "m_62_64" "0.0187636232910056"  "0.0198469631755141"
## m_65_69 "m_65_69" "0.0171713767470092"  "0.0186513629842181"
## m_70_74 "m_70_74" "0.00728192616473006" "0.00860832137733142"
## m_75_99 "m_75_99" "0.00485068426508208" "0.00538020086083214"
## f_16_19 "f_16_19" "0.0173593193561199"  "0.030726924916308"
## f_20_21 "f_20_21" "0.0177796669722571"  "0.0167384026781444"
## f_22_24 "f_22_24" "0.0297245472775364"  "0.0316834050693448"
## f_25_29 "f_25_29" "0.0530516420857978"  "0.0518890483022477"
## f_30_34 "f_30_34" "0.0509260326908187"  "0.0460306073648972"
## f_35_44 "f_35_44" "0.0974323662418713"  "0.100789096126255"
## f_45_54 "f_45_54" "0.10398788731664"    "0.0943328550932568"
## f_55_59 "f_55_59" "0.047537879775097"   "0.0364658058345289"
## f_60_61 "f_60_61" "0.0155768858346592"  "0.0154232424677188"
## f_62_64 "f_62_64" "0.0171231294675505"  "0.0166188426590148"
## f_65_69 "f_65_69" "0.0147765161911754"  "0.0157819225251076"
## f_70_74 "f_70_74" "0.00570390280716364" "0.00502152080344333"
## f_75_99 "f_75_99" "0.00346879745315985" "0.00334768053562889"
```

```r
t.test(nhanes,emp$prop,paired=TRUE)
```

```
##
##  Paired t-test
##
## data:  nhanes and emp$prop
## t = -1.299e-16, df = 25, p-value = 1
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.002379016  0.002379016
```

```
## sample estimates:
## mean of the differences
##          -1.500475e-19
```